LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-498993

Comparative Convergence Analysis of Nonlinear AMLI-cycle MG

X. Hu, P. S. Vassilevski, J. Xu

September 13, 2011

SIAM Journal on Numerical Analysis

Disclaimer

COMPARATIVE CONVERGENCE ANALYSIS OF NONLINEAR AMLI-CYCLE MULTIGRID

XIAOZHE HU, PANAYOT S. VASSILEVSKI, AND JINCHAO XU

Abstract. The main purpose of this paper is to provide a comprehensive convergence analysis of a nonlinear AMLI-cycle multigrid method for symmetric positive definite problems. Based on classical assumptions for approximation and smoothing properties, we show that nonlinear (symmetric and nonsymmetric) AMLI-cycle MG is uniformly convergent. Furthermore, under the only assumption that the smoother is convergent, we show that the nonlinear AMLI-cycle is always better than the respective V-cycle MG. Finally, numerical experiments are presented which illustrate the theoretical results.

1. Introduction

In this paper, we consider the following large-scale sparse linear system of equations

$$Au = f, \tag{1.1}$$

where A is a symmetric positive definite (SPD) operator on a finite-dimensional vector space V. Development of efficient and practical solvers for large sparse linear system of equations arising from discretizations of partial differential equations (PDEs) is an important task in scientific and engineering computing. We consider iterative solution of equation (1.1) by multigrid (MG) methods. MG methods are efficient and often have optimal complexity. There is extensive literature on MG methods; see [12, 26, 27, 5, 9, 22, 28, 25], and references therein for details. MG methods are quite successful in practical applications nowadays. Due to their efficiency and scalability, MG methods, especially their algebraic variants, algebraic multigrid (AMG) methods, have become increasingly used in practice. AMG, originated in [6], gained some popularity after [19] appeared, and more recently have been further extended and developed in various directions ([23, 7, 10, 29, 8, 14], etc.).

In order to improve the robustness of (A)MG methods, we usually use them as preconditioners in Krylov subspace iterative methods [20], such as the conjugate gradient (CG) method in the case when A is SPD.

Date: March 25, 2011–beginning; Today is August 30, 2011.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The third author was supported in part by NSF Grant DMS-0749202 and DMS-0915153, DOE Subcontract B574178 and B591217.

The performance and efficiency of MG methods may degenerate when the physical and geometric properties of the problems become more and more complicated. Generally speaking, if the convergence factor of the two-grid method is too large, the fast convergence property of the MG methods, which is expected to be independent of the levels, can not be guaranteed with standard V- and even W-cycles. A multilevel cycle, which uses the best polynomial approximation of degree n to define the coarse level solver, was originally introduced in [1, 2, 24] and applied to the hierarchical basis MG method. This cycle, which usually is referred to as the algebraic multilevel iteration (AMLI) cycle, is designed to provide optimal condition number, if the degree n of the polynomial is sufficiently large, under the assumption that the V-cycle MG methods have bounded condition number that only depends on the difference of levels. This assumption (on the bounded level length V-cycle convergence) is feasible for certain second order elliptic PDEs without additional assumptions on PDE regularity.

More recently, thanks to the introduction of the nonlinear (variable-step/flexible) preconditioning method and its analysis in [3] (see also [11, 17, 20], etc.), the nonlinear multilevel preconditioners were proposed and their additive version was analyzed in [4]. Furthermore, in [13], the multiplicative version was investigated. In these nonlinear multilevel preconditioners, n steps of a preconditioned CG iterative method replaces the best polynomial approximation and is performed to define the coarse level solvers. The condition number is optimal for properly chosen $n > 1$. The same idea can be adopted to define the MG cycles and has been introduced in [25]. The resulting nonlinear AMLI-cycle MG was analyzed in [18] (see also [25]). In nonlinear AMLI-cycle MG, n steps of a CG method with MG on the coarser level as a preconditioner is applied to define the coarse level solver. Under the assumption that the convergence factor of the V-cycle MG with bounded-level difference is bounded, the uniform convergence property of the nonlinear AMLI-cycle MG methods is shown, if n is chosen to be sufficiently large.

As we can see, the parameter n plays an important rule in the linear and nonlinear AMLI-cycle MG methods. It needs to be sufficiently large to guarantee the uniform convergence even for the problems with full regularity according to the theoretical results. However, one can expect the uniform convergence for these cases for any $n \in \mathbb{Z}^+$, especially $n = 1$, which partly motivated the present work. More specifically, we provide such uniform convergence analysis of the nonlinear AMLI-cycle MG. Under the standard assumptions on approximation and smoothing properties, we show that both nonsymmetric (without post-smoothing) and symmetric (with both pre– and post-smoothing) nonlinear AMLI-cycle MG converge uniformly for any $n \geq 1$ in the following sense:

$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2,$$
$$\|v - \tilde{B}_k[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2,$$

where $\tilde{B}_k^{ns}$ and $\tilde{B}_k$, defined by Algorithm 2.4 and 2.5 below, denote the nonsymmetric and symmetric nonlinear AMLI-cycle MG methods respectively, and the constant $0 < \delta < 1$ is independent of the level k. We also give an alternative proof of the uniform convergence under the assumption used in [18], i.e. the boundedness of V-cycle MG with bounded-level difference. In addition, we show that $\hat{B}_k^{ns}[\cdot]$ and $\hat{B}_k[\cdot]$, the preconditioners used in Krylov subspace iterative methods to define $\tilde{B}_k^{ns}[\cdot]$ and $\tilde{B}_k[\cdot]$, are also uniformly convergent. This means all the recursive calls of Krylov subspace iterative methods can be done only on the coarse levels. On the finest level, we can only do the smoothing steps and still have a uniform convergent method. On the other hand, without the approximation and smoothing properties, similar to MG methods, we are not able to show the uniform convergence for nonlinear AMLI-cycle MG. However, we can compare nonlinear AMLI-cycle MG with V-cycle MG methods, and show that nonlinear AMLI-cycle MG is always better than the corresponding V-cycle MG for any $n \geq 1$. For nonsymmetric case, we can show that

$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k} \leq \|v - B_k^{ns} A_k v\|_{A_k},$$

where B_k^{ns} denotes the nonsymmetric V-cycle MG (without post-smoothing), i.e. \-cycle. For the symmetric case, under the assumption that the smoother is convergent in $\|\cdot\|_{A_k}$ norm, we have

$$\|v - \tilde{B}_k[A_k v]\|_{A_k} \leq \|v - B_k A_k v\|_{A_k},$$

where B_k denotes the V-cycle MG. The above inequality is based on an important property of the full version of nonlinear preconditioned conjugate gradient (PCG) method, namely, the residual of the current iteration is orthogonal to all previous search directions. This property fails for the truncated version of nonlinear PCG method. Therefore, the full version nonlinear PCG should be used rather than the steepest descent method or any truncated version of the nonlinear PCG to define the coarse level solver in the nonlinear AMLI-cycle MG.

The rest of the paper is organized as follows. In section 2, we introduce the nonlinear AMLI-cycle MG algorithms and the basic assumptions. The main results, uniform convergence and comparison theorem of nonlinear AMLI-cycle MG are presented in section 3. In section 4, numerical experiments and the results that illustrate our theoretical results are presented.

2. Preliminaries

Let V be a linear vector space. $(\cdot,\cdot)$ denotes a given inner product on V; its induced norm is $\|\cdot\|$. The adjoint of A with respect to $(\cdot,\cdot)$, denoted by A^t, is defined by $(Au, v) = (u, A^t v)$ for all $u, v \in V$. A is SPD if $A^t = A$ and $(Av, v) > 0$ for all $v \in V \backslash \{0\}$. Since A is SPD with respect to $(\cdot,\cdot)$, $(A\cdot,\cdot)$ defines another inner product on V, denoted by $(\cdot,\cdot)_A$, and its induced norm is $\|\cdot\|_A$.

2.1. **Multigrid.** Let us first introduce the standard V-cycle MG method. Here, we consider the MG methods which are based on a nested sequence of subspaces of V:

$$V_1 \subset V_2 \subset \cdots \subset V_J = V. \tag{2.1}$$

Corresponding to these spaces, we define $Q_k, P_k : V \to V_k$ as the orthogonal projections with respect to $(\cdot,\cdot)$ and $(\cdot,\cdot)_A$ respectively, and define $A_k : V_k \to V_k$ by $(A_k u_k, v_k) = (u_k, v_k)_A$ for $u_k, v_k \in V_k$. Note that A_k is also SPD, and therefore defines a inner product on V_k, denoted by $(\cdot,\cdot)_{A_k}$, and its induced norm is $\|\cdot\|_{A_k}$. We also need to introduce a smoother operator $R_k : V_k \to V_k$ in order to define the multigrid method.

Now we define the nonsymmetric multigrid iterator B_k^{ns} (without post-smoothing) by the following recursive algorithm:

Algorithm 2.1 \-cycle MG: B_k^{ns}

Let $B_1^{ns} = A_1^{-1}$ and assume that $B_{k-1}^{ns} : V_{k-1} \to V_{k-1}$ has been defined, then for $f \in V_k$, $B_k^{ns} : V_k \to V_k$ is defined as follow:

Pre-smoothing: $u_1 = R_k f$;

Coarse grid correction: $B_k^{ns} f := u_1 + B_{k-1}^{ns} Q_{k-1}(f - A_k u_1)$.

Similarly, we can also define the (symmetric) V-cycle multigrid operator B_k recursively:

Algorithm 2.2 V-cycle MG: B_k

Let $B_1 = A_1^{-1}$ and assume that $B_{k-1} : V_{k-1} \to V_{k-1}$ has been defined, then for $f \in V_k$, $B_k : V_k \to V_k$ is defined as follow:

Pre-smoothing $u_1 = R_k f$;

Coarse grid correction $u_2 = u_1 + B_{k-1} Q_{k-1}(f - A_k u_1)$;

Post-smoothing $B_k f := u_2 + R_k^t (f - A_k u_2)$.

2.2. **Nonlinear Preconditioned Conjugate Gradient Method.** In order to introduce nonlinear AMLI-cycle, we also need to introduce the nonlinear PCG Method, which is a simplified version (available for s.p.d. A_k) of the algorithm originated in [3]. The original version in [3] was meant for more general cases including nonsymmetric and possibly indefinite matrices. Let $\hat{B}_k[\cdot] : V_k \to V_k$ be a given nonlinear operator that is intended to approximate the inverse of A_k. We now formulate the nonlinear PCG method that can be used to provide iterated approximate inverse to A_k based on the given nonlinear operator $\hat{B}_k[\cdot]$. This procedure gives another nonlinear operator $\tilde{B}_k[\cdot] : V_k \to V_k$, which can be viewed as an improved approximation to the inverse of A_k.

Algorithm 2.3 Nonlinear PCG Method

Assume we are given a nonlinear operator $\hat{B}_k[\cdot]$ to be used as a preconditioner. Then, for $\forall f \in V_k$, $\tilde{B}_k[f]$ is defined as follows:

Step 1. Let $u_0 = 0$ and $r_0 = f$. Compute $p_0 = \hat{B}_k[r_0]$. Then let

$$u_1 = \alpha_0 p_0, \text{ and } r_1 = r_0 - \alpha_0 A_k p_0, \text{ where } \alpha_0 = \frac{(r_0, p_0)}{(p_0, p_0)_{A_k}}.$$

Step 2. For $i = 1, 2, \cdots, n-1$, compute the next conjugate direction

$$p_i = \hat{B}_k[r_i] + \sum_{j=0}^{i-1} \beta_{i,j} p_j, \text{ where } \beta_{i,j} = -\frac{(\hat{B}_k[r_i], p_j)_{A_k}}{(p_j, p_j)_{A_k}}. \tag{2.2}$$

Then next iterate is

$$u_{i+1} = u_i + \alpha_i p_i, \text{ where } \alpha_i = \frac{(r_i, p_i)}{(p_i, p_i)_{A_k}}, \tag{2.3}$$

and the corresponding residual is

$$r_{i+1} = r_i - \alpha_i A_k p_i. \tag{2.4}$$

Step 3. Let $\tilde{B}_k[f] := u_n$.

Remark 2.1. If we only apply one step of nonlinear PCG method, we can see that

$$\tilde{B}_k[f] = \alpha \hat{B}_k[f], \quad \text{where } \alpha = \frac{(\hat{B}_k[f], f)}{\|\hat{B}_k[f]\|^2_{A_k}}. \tag{2.5}$$

That is, $\tilde{B}_k[f]$ differs from $\hat{B}_k[f]$ by a scalar factor.

Remark 2.2. Due to the choice of $\beta_{i,j}$, it is easy to see that the new direction p_i is orthogonal to all the previous directions p_j, $j = 0, 1, \cdots, i-1$, namely

$$(p_i, p_j)_{A_k} = 0, \ j = 0, 1, 2, \cdots, i-1. \tag{2.6}$$

Due to this property of the directions p_i and the choice of α_i, from (2.4), it is straightforward to see that

$$(r_{i+1}, p_j) = 0, \ j = 0, 1, 2, \cdots, i. \tag{2.7}$$

Finally, by (2.6) and (2.7) , we can show that u_{i+1} computed by (2.3) is the solution of the following minimization problem

$$\min_{\alpha_{i,j} \in \mathbb{R}} \|f - A_k(u_i + \sum_{j=0}^{i} \alpha_{i,j} p_j)\|^2_{A_k^{-1}}.$$

Therefore, we have

$$\|f - A_k u_{i+1}\|^2_{A_k^{-1}} \le \|f - A_k u_i\|^2_{A_k^{-1}},$$

then by induction, we have

$$\|A_k^{-1}f - \tilde{B}_k[f]\|_{A_k}^2 \leq \|A_k^{-1}f - \hat{B}_k[f]\|_{A_k}^2. \tag{2.8}$$

This means, that $\tilde{B}_k[\cdot]$ is a better approximation to A_k^{-1} than $\hat{B}_k[\cdot]$.

Remark 2.3. According to equation (2.2), we use all previous search directions to compute the next one. The resulting Algorithm 2.3 is referred to as the full version of nonlinear PCG method. In practice, due to the memory constraints, we may want to use a truncated version; namely, we only require that the new direction be orthogonal to the $m_i \geq 0$ most recent ones (cf. [17]). In that case, equation (2.2) is replaced by

$$p_i = \hat{B}_k[r_i] + \sum_{j=i-1-m_i}^{i-1} \beta_{i,j} p_j, \text{ where } \beta_{i,j} = -\frac{(\hat{B}_k[r_i], p_j)_{A_k}}{(p_j, p_j)_{A_k}}, \tag{2.9}$$

and the resulting algorithm is called the truncated version of nonlinear PCG method. A general strategy is to have $0 \leq m_i \leq m_{i-1} + 1 \leq i - 1$ and a typical choice is $m_i = 0$. If $p_i = \hat{B}_k[r_i]$ (i.e., formally $m_i = -1$) this choice corresponds to the nonlinear preconditioned steepest descent method. In the present multilevel setting the full version of the method is practically acceptable, since we expect to have relatively few recursive calls (between the levels) and this happens on coarse levels.

If $\hat{B}_k[\cdot]$ approximates the inverse of A_k, with accuracy $\delta \in [0, 1)$, that is,

$$\|A_k^{-1}f - \hat{B}_k[f]\|_{A_k} \leq \delta \|f\|_{A_k^{-1}}. \tag{2.10}$$

We will later use the following convergence results of the nonlinear PCG methods.

Theorem 2.1 (Theorem 10.2, [25])**.** *Assume that $\hat{B}_k[\cdot]$ satisfies* (2.10)*, and $\tilde{B}_k[\cdot]$ is implemented by n iterations of Algorithm 2.3 with $\hat{B}_k[\cdot]$ as the preconditioner, then the following convergence rate estimate holds,*

$$\|A_k^{-1}f - \tilde{B}_k[f]\|_{A_k} \leq \delta^n \|f\|_{A_k^{-1}}. \tag{2.11}$$

Remark 2.4. As stated by Theorem 10.2 in [25], the above convergence rate estimate holds for both full and truncated version of the nonlinear PCG methods.

2.3. **Nonlinear AMLI-cycle MG.** Now, thanks to Algorithm 2.3, we can recursively construct the nonlinear AMLI-cycle MG operator as an approximation of A_k^{-1}. First, we define a nonsymmetric one, i.e. a nonlinear AMLI-cycle MG without post-smoothing.

Similarly to standard (linear) MG, we can also define a symmetric nonlinear AMLI-cycle multigrid by introducing post-smoothing. We have,

Algorithm 2.4 Nonsymmetric nonlinear AMLI-cycle MG: $\hat{B}_k^{ns}[\cdot]$

Assume $\hat{B}_1^{ns}[f] = A_1^{-1}f$, and $\hat{B}_{k-1}^{ns}[\cdot]$ has been defined, then for $f \in V_k$

Pre-smoothing: $u_1 = R_k f$;

Coarse grid correction: $\hat{B}_k^{ns}[f] := u_1 + \tilde{B}_{k-1}^{ns}[Q_{k-1}(f - A_k u_1)]$, where $\tilde{B}_{k-1}^{ns}$ is implemented as in Algorithm 2.3 with $\hat{B}_{k-1}^{ns}$ as preconditioner.

Algorithm 2.5 Nonlinear AMLI-cycle MG: $\hat{B}_k[\cdot]$

Assume $\hat{B}_1[f] = A_1^{-1}f$, and $\hat{B}_{k-1}[\cdot]$ has been defined, then for $f \in V_k$

Pre-smoothing $u_1 = R_k f$;

Coarse grid correction $u_2 = u_1 + \tilde{B}_{k-1}[Q_{k-1}(f - A_k u_1)]$, where $\tilde{B}_{k-1}$ is implemented as in Algorithm 2.3 with $\hat{B}_{k-1}$ as preconditioner;

Post-smoothing $\hat{B}_k[f] := u_2 + R_k^t(f - A_k u_2)$.

2.4. **Assumptions.** Our goal is to analyze the convergence of the nonlinear AMLI-cycle MG using the same assumptions as in the conventional (classical) convergence analysis of MG.

We make the following (classical) assumptions in order to carry out the convergence analysis. These assumptions will be used to show the uniform convergence of the nonlinear AMLI-cycle. The first assumption is the so-called approximation property of the projection P_k.

Assumption 2.1 (Approximation Property)**.**

$$\|(I - P_{k-1})v\|_{A_k}^2 \le \frac{c_1}{\rho(A_k)}\|A_k v\|^2, \quad \forall v \in V_k, \tag{2.12}$$

where $\rho(A_k)$ is the spectral radius of A_k, and c_1 is a constant independent of k.

This assumption is commonly used in the MG literature, for example, Assumption A.7 in [5], "strong approximation property" assumption in [25], and Assumption (A7.1) in [26]. The above assumption holds (see, e.g., [26, 25]) in the case of second order elliptic problems with full regularity.

A next common assumption is on the smoothers. In this paper, we always assume that the (nonsymmetric, in general) smoother R_k, is convergent in the $\|\cdot\|_{A_k}$ norm.

Our second main assumption is that the symmetric composite smoother $\tilde{R}_k$, defined by

$$I - \tilde{R}_k A_k = (I - R_k A_k)(I - R_k^t A_k),$$

satisfies the following smoothing property.

Assumption 2.2 (Smoothing Property)**.**

$$\frac{c_2}{\rho(A_k)}(v, v) \le (\tilde{R}_k v, v), \forall v \in V_k, \tag{2.13}$$

where c_2 is a constant independent of k.

This assumption means that the choice of smoother must be comparable to a simple Richardson smoother. It is used to prove estimates concerning V-cycle MG, see Assumption A.4. in [5]. Note that, we also have another symmetric composite smoother $\bar{R}_k$ which is defined by

$$I - \bar{R}_k A_k = (I - R_k^t A_k)(I - R_k A_k).$$

Based on the two Assumptions 2.1 and 2.2, we have the following (well-known) result (see p. 75 of [26] and p. 145 of [25])

Lemma 2.2. *Assume that Assumptions 2.1 and 2.2 hold, then we have*

$$\|(I - P_{k-1})\hat{v}\|_{A_k}^2 \le \eta(\|v\|_{A_k}^2 - \|\hat{v}\|_{A_k}^2), \tag{2.14}$$

where $\hat{v} = (I - R_k A_k)v$, $v \in V$ *and* $\eta = \dfrac{c_1}{c_2} > 0$ *is a constant independent of* k.

Remark 2.5. The above Lemma can be found as Assumption (A) in [25] and Lemma 6.2 in [26]. It provides perhaps the shortest convergence proof for the V-cycle MG. It is also equivalent to the assumption originally used in [15, 16], see [25] for details. Inequality (2.12) can also be found as inequality (4.82) in [21].

3. Convergence Analysis

In this section, we present the main results of this paper. Based on Assumptions 2.1 and 2.2, we show that nonlinear AMLI-cycle MG is uniformly convergent without the requirement that n, the number of iterations of the nonlinear PCG method, be sufficiently large. Furthermore, without these assumptions, we can also compare nonlinear AMLI-cycle MG with V-cycle MG, and show that the nonlinear AMLI-cycle MG is always better or not worse than the respective V-cycle MG.

The following two representations are useful in our analysis. First, we have a result for the nonsymmetric nonlinear operator $\hat{B}_k^{ns}[\cdot]$ defined in Algorithm 2.4.

Lemma 3.1. *For all* $v \in V_k$

$$v - \hat{B}_k^{ns}[A_k v] = (I - R_k A_k)v - \tilde{B}_{k-1}^{ns}[A_{k-1} P_{k-1}(I - R_k A_k)v], \tag{3.1}$$

and

$$\hat{B}_k^{ns}[v] = R_k v + \tilde{B}_{k-1}^{ns}[Q_{k-1}(I - A_k R_k)v]. \tag{3.2}$$

Proof. Properties (3.1) and (3.2) follow directly from Algorithm 2.4 and the identity $A_{k-1}P_{k-1} = Q_{k-1}A_k$ that holds on V_k. □

Similarly, we have the following lemma concerning the (symmetric) nonlinear operator $\hat{B}_k$ defined in Algorithm 2.5,

Lemma 3.2. *For all $v \in V_k$*

$$v - \hat{B}_k[A_k v] = (I - R_k^t A_k)((I - R_k A_k)v - \tilde{B}_{k-1}[A_{k-1}P_{k-1}(I - R_k A_k)v]), \tag{3.3}$$

and

$$\hat{B}_k[v] = \bar{R}_k v + (I - R_k^t A_k)\tilde{B}_{k-1}[Q_{k-1}(I - A_k R_k)v]. \tag{3.4}$$

Proof. Properties (3.3) and (3.4) are also seen directly from the definition in Algorithm 2.5 (using again the identity $A_{k-1}P_{k-1} = Q_{k-1}A_k$ that holds on V_k). □

3.1. **Uniform Convergence under Assumptions 2.1 and 2.2.** Firstly, we prove the uniform convergence of the nonsymmetric nonlinear AMLI-cycle MG method.

Theorem 3.3. *Let $\hat{B}_k^{ns}[\cdot]$ be defined by Algorithm 2.4, and $\tilde{B}_k^{ns}[\cdot]$ be implemented as in Algorithm 2.3 with $\hat{B}_k^{ns}[\cdot]$ as preconditioner. Assume that Assumptions 2.1 and 2.2 hold, then we have the following uniform estimates*

$$\|v - \hat{B}_k^{ns}[A_k v]\|_{A_k}^2 \le \delta \|v\|_{A_k}^2, \tag{3.5}$$

$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k}^2 \le \delta \|v\|_{A_k}^2, \tag{3.6}$$

where $\delta = \dfrac{c_1}{c_1 + c_2} < 1$, which is a constant independent of k.

Proof. We prove this by mathematical induction. Assume (3.5) and (3.6) hold for $k-1$, and let $\hat{v} = (I - R_k A_k)v$. By Lemma 3.1, we have

$$v - \hat{B}_k^{ns}[A_k v] = \hat{v} - P_{k-1}\hat{v} + P_{k-1}\hat{v} - \tilde{B}_{k-1}^{ns}[A_{k-1}P_{k-1}\hat{v}].$$

Since P_{k-1} is a projection, we have

$$\begin{aligned}
\|v - \hat{B}_k^{ns}[A_k v]\|_{A_k}^2 &= \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \|P_{k-1}\hat{v} - \tilde{B}_{k-1}^{ns}[A_{k-1}P_{k-1}\hat{v}]\|_{A_k}^2 \\
\text{(Induction Assumption)} &\le \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \delta\|P_{k-1}\hat{v}\|_{A_k}^2 \\
\text{(Orthogonality)} &= (1-\delta)\|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \delta\|\hat{v}\|_{A_k}^2 \\
\text{(Lemma 2.2)} &\le (1-\delta)\eta(\|v\|_{A_k}^2 - \|\hat{v}\|_{A_k}^2) + \delta\|\hat{v}\|_{A_k}^2 \\
&= (1-\delta)\eta\|v\|_{A_k}^2 + (\delta - (1-\delta)\eta)\|\hat{v}\|_{A_k}^2 \\
\text{(Choose } \delta = \frac{\eta}{1+\eta} = \frac{c_1}{c_1+c_2}\text{)} &= \delta\|v\|_{A_k}^2.
\end{aligned} \tag{3.7}$$

Then (3.5) follows. Moreover, since $\tilde{B}_k^{ns}[A_k v]$ is obtained by Algorithm 2.3 with $\hat{B}_k^{ns}[\cdot]$ as preconditioner, by (2.8), we have

$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k}^2 \le \|v - \hat{B}_k^{ns}[A_k v]\|_{A_k}^2 \le \delta\|v\|_{A_k}^2. \tag{3.8}$$

This completes the proof. □

In the next theorem we study the convergence of the symmetric nonlinear AMLI-cycle MG under Assumptions 2.1 and 2.2.

Theorem 3.4. *Let $\hat{B}_k[\cdot]$ be defined by Algorithm 2.5, and $\tilde{B}_k^{ns}[\cdot]$ be implemented as in Algorithm 2.3 with $\hat{B}_k[\cdot]$ as the preconditioner. Assume that Assumptions 2.1 and 2.2 hold, then we have the following uniform estimates*

$$\|v - \hat{B}_k[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2, \tag{3.9}$$

$$\|v - \tilde{B}_k[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2, \tag{3.10}$$

where $\delta = \dfrac{c_1}{c_1 + c_2} < 1$ is a constant independent on k.

Proof. Assume that (3.9) and (3.10) hold for $k-1$. Denote $(I - R_k A_k)v$ by $\hat{v}$ as before. Then by Lemma 3.2, we have

$$\begin{aligned}
(v - \hat{B}_k[A_k v], w)_{A_k} &= (\hat{v} - \tilde{B}_{k-1}[A_{k-1} P_{k-1} \hat{v}], \hat{w})_{A_k} \\
&= (\hat{v} - P_{k-1}\hat{v} + P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1} P_{k-1} \hat{v}], \hat{w})_{A_k} \\
&= (\hat{v} - P_{k-1}\hat{v}, \hat{w})_{A_k} + (P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1} P_{k-1} \hat{v}], \hat{w})_{A_k} \\
&= (\hat{v} - P_{k-1}\hat{v}, \hat{w} - P_{k-1}\hat{w})_{A_k} + (P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1} P_{k-1} \hat{v}], P_{k-1}\hat{w})_{A_k} \\
\text{(Cauchy Schwarz)} &\leq \|\hat{v} - P_{k-1}\hat{v}\|_{A_k} \|\hat{w} - P_{k-1}\hat{w}\|_{A_k} \\
&\quad + \|P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1} P_{k-1} \hat{v}]\|_{A_k} \|P_{k-1}\hat{w}\|_{A_k} \\
\text{(Induction Assumption)} &\leq \|\hat{v} - P_{k-1}\hat{v}\|_{A_k} \|\hat{w} - P_{k-1}\hat{w}\|_{A_k} + \delta^{1/2} \|P_{k-1}\hat{v}\|_{A_k} \|P_{k-1}\hat{w}\|_{A_k} \\
\text{(Cauchy Schwarz)} &\leq \sqrt{\|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \delta \|P_{k-1}\hat{v}\|_{A_k}^2} \times \sqrt{\|\hat{w} - P_{k-1}\hat{w}\|_{A_k}^2 + \|P_{k-1}\hat{w}\|_{A_k}^2}
\end{aligned}$$

For the first term, we can use the same argument in Theorem 3.4. The second term can be estimated using the orthogonality of P_{k-1} and the assumption that the smoother R_k is A_k-convergent. Therefore, we have

$$(v - \hat{B}_k[A_k v], w)_{A_k} \leq \delta^{1/2} \|v\|_{A_k} \|w\|_{A_k}, \qquad \delta = \eta/(\eta + 1).$$

This shows that $\|v - \hat{B}_k[A_k v]\|_{A_k} \leq \delta^{1/2} \ \|v\|_{A_k}$ with $\delta = \eta/(\eta+1)$, and hence (3.9) follows by mathematical induction.

Note that $\tilde{B}_k[A_k v]$ is obtained by Algorithm 2.3 with $\hat{B}_k[\cdot]$ used as preconditioner, hence by (2.8), we have

$$\|v - \tilde{B}_k[A_k v]\|_{A_k}^2 \leq \|v - \hat{B}_k[A_k v]\|_{A_k}^2,$$

Then (3.10) follows directly. □

In Theorem 3.3 and 3.4, the full version of the nonlinear PCG was (implicitly) assumed. However, it is clear that since we only use the minimization property (2.8) in the proof, the final result also holds for any truncated version of the nonlinear PCG. Therefore, we have the following two corollaries regarding the uniform convergence of the nonlinear AMLI-cycle MG using truncated versions of the nonlinear PCG.

Corollary 3.5. *Let $\hat{B}_k^{ns}[\cdot]$ be defined by Algorithm 2.4, and $\tilde{B}_k^{ns}[\cdot]$ be implemented in a truncated version of Algorithm 2.3 with $\hat{B}_k^{ns}[\cdot]$ as preconditioner. Assume that Assumptions 2.1 and 2.2 hold, then we have the following uniform estimates*

$$\|v - \hat{B}_k^{ns}[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2, \tag{3.11}$$

$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2, \tag{3.12}$$

where $\delta = \dfrac{c_1}{c_1 + c_2} < 1$ is a constant independent on k.

Corollary 3.6. *Let $\hat{B}_k[\cdot]$ be defined by Algorithm 2.5, and $\tilde{B}_k^{ns}[\cdot]$ be implemented in a truncated version of Algorithm 2.3 with $\hat{B}_k[\cdot]$ as preconditioner. Assume that Assumptions 2.1 and 2.2 hold, then we have the following uniform estimates*

$$\|v - \hat{B}_k[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2, \tag{3.13}$$

$$\|v - \tilde{B}_k[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2, \tag{3.14}$$

where $\delta = \dfrac{c_1}{c_1 + c_2} < 1$ is a constant independent on k.

Remark 3.1. In [18], uniform convergence of nonlinear AMLI-cycle MG is shown if the number of nonlinear PCG iterations is chosen to be sufficiently large (under certain assumption on the boundedness of the V-cycle MG with bounded-level difference). However, this condition is not needed in the above theorems. Our uniform convergence results hold for arbitrary choice of the number of nonlinear PCG iterations but requires instead Assumptions 2.1 and 2.2.

3.2. **Convergence results without Assumptions 2.1 and 2.2.** So far, the convergence results suggest that not only $\tilde{B}_k[\cdot]$ but also $\hat{B}_k[\cdot]$ converge uniformly under Assumption 2.1 and 2.2. A natural question arises, under the same assumption on the bounded convergence factor of the V-cycle MG with bounded-level difference, k_0, used in [18], does the nonlinear operator $\hat{B}_k[\cdot]$ converge uniformly when n is sufficiently large? The following two theorems give a positive answer to this question. This is a (slight) generalization of the result in [18].

For the sake of simplicity, let us assume that the convergence factor of two-grid method ($k_0 = 1$) is independent of k. The more general case, when the convergence factor of V-cycle MG with bounded-level difference k_0 is independent of k, can be analyzed similarly.

Theorem 3.7. *Let $\hat{B}_k^{ns}[\cdot]$ be defined by Algorithm 2.4, and $\tilde{B}_k^{ns}[\cdot]$ be implemented as in Algorithm 2.3 with $\hat{B}_k^{ns}[\cdot]$ as preconditioner. Assume that the convergence factor of two-grid method is bounded by $\bar{\delta} \in [0,1)$ which is independent of k, Let n, the number of iterations of the nonlinear PCG method, be chosen such that the following inequality*

$$(1-\delta^n)\bar{\delta} + \delta^n \le \delta, \tag{3.15}$$

has a solution $\delta \in [0,1)$. A sufficient condition for this is

$$n > \frac{1}{1-\bar{\delta}}. \tag{3.16}$$

then we have the following uniform estimates

$$\|v - \hat{B}_k^{ns}[A_k v]\|_{A_k}^2 \le \delta \|v\|_{A_k}^2, \tag{3.17}$$

$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k}^2 \le \delta^n \|v\|_{A_k}^2, \tag{3.18}$$

where δ is independent of k.

Proof. We prove the estimates by mathematical induction. Assume that (3.17) and (3.18) hold for $k-1$, and let $\hat{v} = (I - R_k A_k)v$. Similar to Theorem 3.3, we have

$$\begin{aligned}
\|v - \hat{B}_k^{ns}[A_k v]\|_{A_k}^2 &= \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \|P_{k-1}\hat{v} - \tilde{B}_{k-1}^{ns}[A_{k-1}P_{k-1}\hat{v}]\|_{A_k}^2 \\
\text{(Induction Assumption)} &\le \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \delta^n \|P_{k-1}\hat{v}\|_{A_k}^2 \\
\text{(orthogonality)} &= (1-\delta^n)\|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \delta^n \|\hat{v}\|_{A_k}^2 \\
\text{(two-grid convergence factor is bounded)} &\le ((1-\delta^n)\bar{\delta} + \delta^n)\|v\|_{A_k}^2 \\
\text{(by (3.15))} &\le \delta \|v\|_{A_k}^2.
\end{aligned}$$

This shows that estimate (3.17) holds. Moreover, According to Theorem 2.1, and letting $f = A_k v$ in (2.11), the estimate (3.18) follows directly.

Now we show that (3.16) implies that there exists a δ which solves (3.15). (3.15) is equivalently to

$$\phi(\delta) \equiv (1 + \delta + \delta^2 + \cdots + \delta^{n-1})\bar{\delta} - (\delta + \delta^2 + \cdots + \delta^{n-1}) \le 0.$$

Due to (3.16), $\phi(1) = n\bar{\delta} - (n-1) = 1 - n(1-\bar{\delta}) < 0$, and $\phi(0) = \bar{\delta} > 0$. therefore, there is a $\delta^* \in [0,1)$ such that $\phi(\delta^*) = 0$. Then any $\delta \in [\delta^*, 1)$ will satisfy (3.15). □

Similar result also holds for symmetric case, see the following theorem.

Theorem 3.8. *Let $\hat{B}_k[\cdot]$ be defined by Algorithm 2.5, and $\tilde{B}_k[\cdot]$ be implemented as in Algorithm 2.3 with $\hat{B}_k[\cdot]$ as preconditioner. Assume that the convergence factor of two-grid method is bounded by $\bar{\delta} \in [0,1)$ which is independent of k, Let n, the number of iterations of the nonlinear PCG method, be chosen such that the following inequality*

$$(1-\delta^n)\bar{\delta} + \delta^n \le \delta,$$

has a solution $\delta \in [0,1)$. A sufficient condition for this is

$$n > \frac{1}{1-\bar{\delta}}.$$

Then we have the following uniform estimates

$$\|v - \hat{B}_k[A_k v]\|_{A_k}^2 \leq \delta \|v\|_{A_k}^2, \tag{3.19}$$

$$\|v - \tilde{B}_k[A_k v]\|_{A_k}^2 \leq \delta^n \|v\|_{A_k}^2, \tag{3.20}$$

where δ is independent of k.

Proof. Assume that (3.19) and (3.20) hold for $k-1$. Similar to Theorem 3.4, we have

$$\begin{aligned}
(v - \hat{B}_k[A_k v], w)_{A_k} &= (\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], \hat{w})_{A_k} \\
\text{(Cauchy Schwarz)} &\leq \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}\|\hat{w} - P_{k-1}\hat{w}\|_{A_k} \\
&\quad + \|P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}]\|_{A_k}\|P_{k-1}\hat{w}\|_{A_k} \\
\text{(Induction Assumption)} &\leq \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}\|\hat{w} - P_{k-1}\hat{w}\|_{A_k} + \delta^{n/2}\|P_{k-1}\hat{v}\|_{A_k}\|P_{k-1}\hat{w}\|_{A_k} \\
\text{(Cauchy Schwarz)} &\leq \sqrt{\|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \delta^n\|P_{k-1}\hat{v}\|_{A_k}^2} \times \sqrt{\|\hat{w} - P_{k-1}\hat{w}\|_{A_k}^2 + \|P_{k-1}\hat{w}\|_{A_k}^2}.
\end{aligned}$$

The first term on the right hand side can be estimated by the same argument as in Theorem 3.7, therefore, we have

$$(v - \hat{B}_k[A_k v], w)_{A_k} \leq \delta^{1/2}\|v\|_{A_k}\|w\|_{A_k},$$

this implies (3.19). Moreover, According to Theorem 2.1, and letting $f = A_k v$ in (2.11), the estimate (3.20) follows directly.

The existence of δ has been shown in Theorem 3.7. □

For $k_0 = 1$ and $n = 2$, the nonlinear AMLI-cycle MG has the complexity of W-cycle MG, and the sufficient condition (3.16) becomes

$$2 = n > \frac{1}{1-\bar{\delta}} \Rightarrow \bar{\delta} < \frac{1}{2}.$$

In conclusion, we have the following result.

Corollary 3.9. *If the two-grid method at any level k (with exact solution at coarse level $k+1$) has a uniformly bounded convergence rate $\bar{\delta} < \frac{1}{2}$, then the respective nonlinear AMLI-cycle MG with $n = 2$ converges uniformly.*

Remark 3.2. Since Theorem 2.1 holds for both full and truncated version of the nonlinear PCG methods, the above uniform convergence estimates also hold for both full and truncated version of the nonlinear PCG methods.

3.3. **Comparison Analysis.** Without Assumptions 2.1 and 2.2, although we cannot show that the nonlinear AMLI-cycle MG is uniformly convergent, we can compare it with the nonlinear AMLI-cycle MG with the corresponding nonsymmetric (\-cycle) and symmetric (V-cycle) MG. In this section, we show that the nonlinear AMLI-cycle is always better (or not worse) under the assumption that smoother is convergent in the $\|\cdot\|_{A_k}$-norm.

The first comparison theorem concerns with the nonsymmetric nonlinear AMLI-cycle MG and \-cycle MG. It shows that both the nonlinear operator $\hat{B}_k^{ns}$ and $\tilde{B}_k^{ns}$ give better approximations to the inverse of A_k.

Theorem 3.10. *Let B_k^{ns} and $\hat{B}_k^{ns}[\cdot]$ be defined by Algorithm 2.1 and 2.4 , and $\tilde{B}_k^{ns}[\cdot]$ be implemented as in Algorithm 2.3 with $\hat{B}_k^{ns}[\cdot]$ used as preconditioner. Then we have*

$$\|v - \hat{B}_k^{ns}[A_k v]\|_{A_k} \leq \|v - B_k^{ns} A_k v\|_{A_k}. \tag{3.21}$$

$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k} \leq \|v - B_k^{ns} A_k v\|_{A_k}. \tag{3.22}$$

Proof. We use mathematical induction to prove the theorem. Assume that (3.21) and (3.22) hold for $k-1$. By Algorithm 2.1, we have

$$(I - B_k^{ns} A_k) v = \hat{v} - B_{k-1}^{ns} A_{k-1} P_{k-1} \hat{v} = \hat{v} - P_{k-1}\hat{v} + P_{k-1}\hat{v} - B_{k-1}^{ns} A_{k-1} P_{k-1} \hat{v}, \tag{3.23}$$

where $\hat{v} = (I - R_k A_k) v$ as before. Note that P_{k-1} is a projection, hence we have

$$\|v - B_k^{ns} A_k v\|_{A_k}^2 = \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \|P_{k-1}\hat{v} - B_{k-1}^{ns} A_{k-1} P_{k-1}\hat{v}\|_{A_k}^2 \tag{3.24}$$

Similarly, for the nonlinear operator $\hat{B}_k^{ns}[\cdot]$, by Lemma 3.1, we have

$$\begin{aligned} \|v - \hat{B}_k^{ns}[A_k v]\|_{A_k}^2 &= \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \|P_{k-1}\hat{v} - \tilde{B}_{k-1}^{ns}[A_{k-1} P_{k-1}\hat{v}]\|_{A_k}^2 \\ &\leq \|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \|P_{k-1}\hat{v} - B_{k-1}^{ns} A_{k-1} P_{k-1}\hat{v}\|_{A_k}^2 \\ &= \|v - B_k^{ns} A_k v\|_{A_k}^2 \end{aligned} \tag{3.25}$$

Then (3.21) follows. Moreover, since $\tilde{B}_k^{ns}[A_k v]$ is obtained by Algorithm 2.3 with $\hat{B}_k^{ns}[\cdot]$ used as preconditioner, by (2.8), we have

$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k}^2 \leq \|v - \hat{B}_k^{ns}[A_k v]\|_{A_k}^2 \leq \|v - B_k^{ns} A_k v\|_{A_k}^2. \tag{3.26}$$

This completes the proof. $\square$

As before, we only used the minimization property (2.8) in the proof, therefore, Theorem 3.10 is also true when we use any truncated version of the nonlinear PCG method to define the coarse level solver. Thus, we have the following corollary.

Corollary 3.11. *Let B_k^{ns} and $\hat{B}_k^{ns}[\cdot]$ be defined by Algorithm 2.1 and 2.4. Let also $\tilde{B}_k^{ns}[\cdot]$ be implemented as in a truncated version of Algorithm 2.3 with $\hat{B}_k^{ns}[\cdot]$ as preconditioner. Then,* *we have*

$$\|v - \hat{B}_k^{ns}[A_k v]\|_{A_k} \le \|v - B_k^{ns} A_k v\|_{A_k},$$
$$\|v - \tilde{B}_k^{ns}[A_k v]\|_{A_k} \le \|v - B_k^{ns} A_k v\|_{A_k}.$$

Next we show that, similarly to the nonsymmetric case, the nonlinear AMLI-cycle is better (not worse) than the respective V-cycle MG, and both nonlinear operator $\hat{B}_k$ and $\tilde{B}_k$ provide better approximations to the inverse of A_k.

We first show the following key property of the nonlinear operator $\tilde{B}_k[\cdot]$ obtained by Algorithm 2.3. This property plays an important rule in our analysis.

Lemma 3.12. *Let $\tilde{B}_k[\cdot]$ be implemented as in Algorithm 2.3 with $\hat{B}_k[\cdot]$ as preconditioner. For $\forall v \in V_k$, we have*

$$\|v - \tilde{B}_k[A_k v]\|_{A_k}^2 = (v - \tilde{B}_k[A_k v], v)_{A_k}. \tag{3.27}$$

Proof. By (2.3), we can see that $u_i = \sum_{j=0}^{i-1} \alpha_j p_j$. Due to the fact that the residual r_i is orthogonal to all the previous directions p_j, $j = 0, 1, \cdots, i-1$, we have $(r_i, u_i) = 0$. By definition, $\tilde{B}_k[f] := u_n$, hence we have $(r_n, u_n) = 0$, $r_n = f - A_k \tilde{B}_k[f]$, i.e.,

$$(f - A_k \tilde{B}_k[f], \tilde{B}_k[f]) = 0. \tag{3.28}$$

Letting $f = A_k v$, we have

$$\begin{aligned}\|v - \tilde{B}_k[A_k v]\|_{A_k}^2 &= (v - \tilde{B}_k[A_k v], v - \tilde{B}_k[A_k v])_{A_k} \\ &= (v - \tilde{B}_k[A_k v], v)_{A_k} + (v - \tilde{B}_k[A_k v], \tilde{B}_k[A_k v])_{A_k}.\end{aligned}$$

The second term vanishes due to (3.28) and the choice $f = A_k v$. Then (3.27) follows directly. □

Now we are in a position to show the following comparison theorem for the nonlinear AMLI-cycle MG and the respective V-cycle MG.

Theorem 3.13. *Let $\hat{B}_k[\cdot]$ be defined by Algorithm 2.5, and $\tilde{B}_k[\cdot]$ be implemented as in Algorithm 2.3 with $\hat{B}_k[\cdot]$ as preconditioner. We also assume that the smoother R_k is convergent. For $\forall v \in V_k$, the following estimates hold:*

$$0 \le (v - \tilde{B}_k[A_k v], v)_{A_k} \le (v - \hat{B}_k[A_k v], v)_{A_k} \le (v - B_k A_k v, v)_{A_k}. \tag{3.29}$$

Proof. Inequalities (3.29) hold trivially for $k = 1$. Assuming by induction that (3.29) holds for $k-1$, by Lemma 3.12, we than have that

$$(v - \tilde{B}_k[A_k v], v)_{A_k} = \|v - \tilde{B}_k[A_k v]\|_{A_k}^2 \ge 0,$$

which confirms the first inequality in (3.29). Since $\tilde{B}_k[A_k v]$ is obtained by Algorithm 2.3 with $\hat{B}_k[\cdot]$ as preconditioner, by (2.8), we have

$$(v - \tilde{B}_k[A_k v], v)_{A_k} = \|v - \tilde{B}_k[A_k v]\|^2_{A_k} \le \|v - \hat{B}_k[A_k v]\|^2_{A_k}.$$

On the other hand, letting $\hat{v} = (I - R_k A_k)v$, according to Lemma 3.2, we have

$$\begin{aligned}
\|v - \hat{B}_k[A_k v]\|^2_{A_k} &= \|(I - R_k^t A_k)(\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}])\|^2_{A_k} \\
\text{(smoother is convergent)} &\le \|\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}]\|^2_{A_k} \\
&= \|\hat{v} - P_{k-1}\hat{v} + P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}]\|^2_{A_k} \\
\text{(orthogonality)} &= \|\hat{v} - P_{k-1}\hat{v}\|^2_{A_k} + \|P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}]\|^2_{A_k} \\
&= \|\hat{v} - P_{k-1}\hat{v}\|^2_{A_k} + \|P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}]\|^2_{A_{k-1}} \\
\text{(Lemma 3.12)} &= (\hat{v} - P_{k-1}\hat{v}, \hat{v} - P_{k-1}\hat{v})_{A_k} + (P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], P_{k-1}\hat{v})_{A_k} \\
\text{(orthogonality)} &= (\hat{v} - P_{k-1}\hat{v}, \hat{v})_{A_k} + (P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], \hat{v})_{A_k} \\
&= (\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], \hat{v})_{A_k} \\
&= (v - \hat{B}_k[A_k v], v)_{A_k}.
\end{aligned}$$

Therefore, we have

$$(v - \tilde{B}_k[A_k v], v)_{A_k} \le (v - \hat{B}_k[A_k v], v)_{A_k},$$

which confirms the second inequality in (3.29). For the last inequality, we have

$$\begin{aligned}
(v - \hat{B}_k[A_k v], v)_{A_k} &= (\hat{v} - P_{k-1}\hat{v}, \hat{v})_{A_k} + (P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], \hat{v})_{A_k} \\
&= (\hat{v} - P_{k-1}\hat{v}, \hat{v})_{A_k} + (P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], P_{k-1}\hat{v})_{A_k} \\
\text{(induction assumption)} &\le (\hat{v} - P_{k-1}\hat{v}, \hat{v})_{A_k} + (P_{k-1}\hat{v} - B_{k-1}A_{k-1}P_{k-1}\hat{v}, P_{k-1}\hat{v})_{A_k} \\
&= (\hat{v} - P_{k-1}\hat{v}, \hat{v})_{A_k} + (P_{k-1}\hat{v} - B_{k-1}A_{k-1}P_{k-1}\hat{v}, \hat{v})_{A_k} \\
&= (\hat{v} - B_{k-1}A_{k-1}P_{k-1}\hat{v}, \hat{v})_{A_k} \\
&= (v - B_k A_k v, v)_{A_k}.
\end{aligned}$$

This confirms the last inequality in (3.29) and thus the proof is complete. □

Remark 3.3. We recall that Lemma 3.12 is based on the fact that the current residual r_i is orthogonal to all previous search direction, which only holds for the full version of the nonlinear AMLI-cycle MG. Therefore, the full version of the nonlinear PCG should be preferred in practice than the steepest descent method or truncated version of the nonlinear PCG, since then we have guaranteed monotonicity as stated in Theorem 3.13 (which holds only if the full version of the nonlinear PCG method is applied).

3.4. **Comparison results under Assumptions 2.1 and 2.2.** We return now to the Assumption 2.1 and 2.2. Under these assumptions, we have the following comparison theorem which shows that the nonlinear AMLI-cycle MG is (strictly) better than \-cycle MG with a factor $\rho < 1$, which we specify in the following theorem.

Theorem 3.14. *Let $\hat{B}_k[\cdot]$ be defined by Algorithm 2.5, and $\tilde{B}_k[\cdot]$ be implemented as in Algorithm 2.3 with $\hat{B}_k[\cdot]$ as a preconditioner. Assume that Assumption 2.1 and 2.2 hold. We then have the estimates*

$$\|v - \hat{B}_k[A_k v]\|_{A_k} \le \rho \|v - B_k^{ns} A_k v\|_{A_k} \tag{3.30}$$

$$\|v - \tilde{B}_k[A_k v]\|_{A_k} \le \rho \|v - B_k^{ns} A_k v\|_{A_k}, \tag{3.31}$$

where $\rho = \sqrt{\dfrac{c_1}{c_1 + c_2}} < 1$ which is a constant independent of k.

Proof. Assume that (3.30) and (3.31) hold for $k-1$. Denote $(I - R_k A_k)v$ by $\hat{v}$ as before. Then by Lemma 3.2, we have

$$\begin{aligned}
(v - \hat{B}_k[A_k v], w)_{A_k} &= (\hat{v} - \tilde{B}_{k-1}[A_{k-1} P_{k-1}\hat{v}], \hat{w})_{A_k} \\
&= (\hat{v} - P_{k-1}\hat{v} + P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], \hat{w})_{A_k} \\
&= (\hat{v} - P_{k-1}\hat{v}, \hat{w})_{A_k} + (P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], \hat{w})_{A_k} \\
&= (\hat{v} - P_{k-1}\hat{v}, \hat{w} - P_{k-1}\hat{w})_{A_k} + (P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}], P_{k-1}\hat{w})_{A_k} \\
\text{(Cauchy Schwarz)} &\le \|\hat{v} - P_{k-1}\hat{v}\|_{A_k} \|\hat{w} - P_{k-1}\hat{w}\|_{A_k} \\
&\quad + \|P_{k-1}\hat{v} - \tilde{B}_{k-1}[A_{k-1}P_{k-1}\hat{v}]\|_{A_k} \|P_{k-1}\hat{w}\|_{A_k} \\
\text{(induction assumption)} &\le \|\hat{v} - P_{k-1}\hat{v}\|_{A_k} \|\hat{w} - P_{k-1}\hat{w}\|_{A_k} \\
&\quad + \rho \|P_{k-1}\hat{v} - \tilde{B}_{k-1}^{ns} A_{k-1} P_{k-1}\hat{v}\|_{A_k} \|P_{k-1}\hat{w}\|_{A_k} \\
\text{(Cauchy Schwarz)} &\le \sqrt{\|\hat{v} - P_{k-1}\hat{v}\|_{A_k}^2 + \|P_{k-1}\hat{v} - B_{k-1}^{ns} A_{k-1} P_{k-1}\hat{v}\|_{A_k}^2} \\
&\quad \times \sqrt{\|\hat{w} - P_{k-1}\hat{w}\|_{A_k}^2 + \rho^2 \|P_{k-1}\hat{w}\|_{A_k}^2} \\
&= \|v - B_k^{ns} A_k v\|_{A_k} \times \sqrt{\|\hat{w} - P_{k-1}\hat{w}\|_{A_k}^2 + \rho^2 \|P_{k-1}\hat{w}\|_{A_k}^2}
\end{aligned}$$

Note that

$$\begin{aligned}
\|\hat{w} - P_{k-1}\hat{w}\|_{A_k}^2 + \rho^2 \|P_{k-1}\hat{w}\|_{A_k}^2 &= (1 - \rho^2)\|\hat{w} - P_{k-1}\hat{w}\|_{A_k}^2 + \rho^2 \|\hat{w}\|_{A_k}^2 \\
\text{(Lemma 2.2)} &\le (1 - \rho^2)\eta(\|w\|_{A_k}^2 - \|\hat{w}\|_{A_k}^2) + \rho^2 \|\hat{w}\|_{A_k}^2 \\
&= (1 - \rho^2)\eta \|w\|_{A_k}^2 + (\rho^2 - (1 - \rho^2)\eta)\|\hat{w}\|_{A_k}^2 \\
&= \rho^2 \|w\|_{A_k}^2 \quad \text{(choose } \rho^2 = \frac{\eta}{1 + \eta}\text{)}.
\end{aligned}$$

Therefore, we have

$$(v - \hat{B}_k[A_k v], w)_{A_k} \le \|v - B_k^{ns} A_k v\|_{A_k} \times \rho \|w\|_{A_k},$$

which implies (3.30). Moreover, since $\tilde{B}_k[A_k v]$ is obtained by Algorithm 2.3 with $\hat{B}_k[\cdot]$ as the preconditioner, by (2.8), we have

$$\|v - \tilde{B}_k[A_k v]\|_{A_k}^2 \le \|v - \hat{B}_k[A_k v]\|_{A_k}^2.$$

Then (3.31) follows from the proven estimate (3.30). □

4. Numerical Experiments

In this section, we present some numerical results to illustrate our theoretical results. The first model problem we consider here is

$$-\Delta u = f, \quad \text{in } \Omega, \tag{4.1}$$

$$u = 0, \quad \text{on } \partial\Omega, \tag{4.2}$$

where Ω is the unit square in $\mathbb{R}^2$. In our numerical experiments, we discretize equation (1.1) by linear finite element method with the choice of $f = 1$. The domain Ω is triangulated by uniform refinements and the mesh size on the finest level is $h = 2^{-k}$, where k is the number of levels used.

In Table 4.1, the numerical results of nonlinear AMLI-cycle MG and V-cycle MG methods are presented and compared. Under the setting of our experiments, Assumption 2.1 and 2.2 are satisfied, then according to Theorem 3.4, both nonlinear operator $\hat{B}_k[\cdot]$ and $\tilde{B}_k[\cdot]$ are uniformly convergent, which is illustrated with the numerical results shown in Table 4.1. Furthermore, we can see that $\hat{B}_k$ and $\tilde{B}_k$ are better than B_k in terms of the number of iterations, which agrees with Theorem 3.13.

The second model problem is a diffusion equation with large jump in the coefficient

$$-\nabla \cdot (a(x)\nabla u) = f, \quad \text{in } \Omega, \tag{4.3}$$

$$u = 0, \quad \text{on } \partial\Omega, \tag{4.4}$$

where $\Omega = (0,1)\times(0,1)$. We have $a(x) = 1$ on $\Omega_1 = (0.25, 0.5)\times(0.25, 0.5)$ and $\Omega_2 = (0.5, 0.75)\times(0.5, 0.75)$, and $a(x) = 10^{-6}$ on $\Omega\backslash(\bar{\Omega}_1 \cup \bar{\Omega}_2)$. The domain Ω is triangulated by uniform refinements and the mesh size on the finest level is $h = 2^{-k}$, where k is the number of levels used. In this test problem, we choose $f = 0$, which means the exact solution is $u^* = 0$. Since we know the exact solution, the stopping criteria is $||u^* - u_i||_A \le 10^{-6}$, where u_i is the i-th iteration of the MG methods.

It is well known that the performance of V-cycle MG methods for this jump coefficient problem will degenerate. Table 4.2 confirms this fact. For this problem, due to lack of regularity, if one iteration of nonlinear PCG methods is used to define the coarse level solver,

Table 4.1. Number of iterations of the V-cycle MG and nonlinear AMLI-cycle MG. (stopping criteria: relative residual is less than 10^{-6}; N-PCG(n): n iterations of the nonlinear PCG is used to define the coarse level solver $\tilde{B}_{k-1}[\cdot]$)

	B_k	$\hat{B}_k[\cdot]$		$\tilde{B}_k[\cdot]$	
k		N-PCG(1)	N-PCG(2)	N-PCG(1)	N-PCG(2)
5	9	9	9	7	3
6	11	10	10	8	4
7	12	11	10	9	4
8	13	11	10	10	4
9	13	12	10	10	4
10	14	12	10	11	4
11	14	12	10	12	4
12	14	13	10	12	4

both $\hat{B}_k[\cdot]$ and $\tilde{B}_k[\cdot]$ appear to be non–uniformly convergent. Nevertheless, according to Theorem 3.13, they exhibit better convergence than the V-cycle MG. Furthermore, if the number of iterations of the nonlinear PCG methods is sufficiently large ($n = 2$ in this case), according to the theoretical results in [18], we can expect that $\tilde{B}_k[\cdot]$ be uniformly convergent both with respect to the number of levels k and the jumps, which is demonstrated by the numerical results shown in Table 4.2. Furthermore, we see that $\hat{B}_k[\cdot]$ also converges uniformly.

Table 4.2. Number of iterations of the V-cycle MG and nonlinear AMLI-cycle MG for jump coefficient problem. (stopping criteria: energy norm of error is less than 10^{-6}; N-PCG(n): n iterations of the nonlinear PCG is used to define the coarse level solver $\tilde{B}_{k-1}[\cdot]$)

	B_k	$\hat{B}_k[\cdot]$		$\tilde{B}_k[\cdot]$	
k		N-PCG(1)	N-PCG(2)	N-PCG(1)	N-PCG(2)
5	27	15	13	13	4
6	40	22	14	15	5
7	49	29	14	20	5
8	56	37	15	30	5
9	76	45	15	42	5
10	102	55	15	47	5

In the last numerical experiment, we use unsmoothed aggregation AMG (UA-AMG) methods to solve model problem (4.1) discretized by linear finite element on uniform meshes. Given the k-th level matrix $A_k \in \mathbb{R}^{n_k \times n_k}$, in the UA-AMG method we define prolongation matrix

P_{k-1}^k from a non-overlapping partition of the n_k unknowns at level k into n_{k-1} nonempty disjoint sets G_j, $j = 1, \ldots, n_{k-1}$, which are referred to as aggregates. In our numerical experiments, we use Algorithm 2 in [23] to generate the aggregates on each level. Once the aggregates are constructed, the prolongator P_{k-1}^k is a $n_k \times n_{k-1}$ matrix given by

$$(P_{k-1}^k)_{ij} = \begin{cases} 1 & \text{if } i \in G_j \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \ldots, n_k, \quad j = 1, \ldots, n_{k-1}.$$

With such piecewise constant prolongation, the coarse level matrix $A_{k-1} \in \mathbb{R}^{n_{k-1} \times n_{k-1}}$ is defined by

$$A_{k-1} = (P_{k-1}^k)^t A_k (P_{k-1}^k).$$

Since now we consider AMG methods, we do not have the orthogonal projections Q_k and P_k, and cannot use them to define the operators B_k, $\hat{B}_k[\cdot]$ and $\tilde{B}_k[\cdot]$. However, thanks to the prolongation P_{k-1}^k, the V-cycle MG iterator B_k for UA-AMG is defined recursively by Algorithm 2.2 with the following coarse grid correction step,

$$u_2 = u_1 + P_{k-1}^k B_{k-1} (P_{k-1}^k)^t (f - A_k u_1).$$

Similarly, the nonlinear operator $\hat{B}_k[\cdot]$ for UA-AMG is defined by Algorithm 2.5 with the following coarse grid correction step

$$u_2 = u_1 + P_{k-1}^k \tilde{B}_{k-1}[(P_{k-1}^k)^t (f - A_k u_1)],$$

and the nonlinear operator $\tilde{B}_k[\cdot]$ for UA-AMG is implemented as in Algorithm 2.3 with $\hat{B}_k[\cdot]$ for UA-AMG as preconditioner.

The results are shown in Table 4.3. We can see that if we use V-cycle MG for UA-AMG, the number of iterations depends strongly on the size of the problem. If we use one iteration of the nonlinear PCG to define the coarse level solver, the performance of $\hat{B}_k[\cdot]$ and $\tilde{B}_k[\cdot]$ still depends on the size of the problem, but the number of iterations grows much slower. If we use two iterations, both $\hat{B}_k[\cdot]$ and $\tilde{B}_k[\cdot]$ exhibit uniform convergence.

The last experiments demonstrate the potential of the nonlinear AMLI-cycle MG methods in cases when the constructed hierarchy of interpolation matrices is not energy stable. In many cases it is straightforward to come up with simple (e.g., block–diagonal) interpolation matrices which however lead to V-cycle MG that generally has level–dependent convergence. The nonlinear AMLI-cycle can be used in such instances to substantially improve the convergence (cf., e.g., [14]).

References

[1] O. Axelsson and P. S. Vassilevski. Algebraic multilevel preconditioning methods. I. *Numer. Math.*, 56(2-3):157–177, 1989.

Table 4.3. Number of iterations of the V-cycle MG and nonlinear AMLI-cycle MG for the UA-AMG method. (stopping criteria: relative residual is less than 10^{-6}; N-PCG(n): n iterations of the nonlinear PCG is used to define the coarse level solver $\tilde{B}_{k-1}[\cdot]$)

	B_k	$\hat{B}_k[\cdot]$		$\tilde{B}_k[\cdot]$	
Size		N-PCG(1)	N-PCG(2)	N-PCG(1)	N-PCG(2)
3,969	100	48	40	34	9
16,129	244	70	41	38	9
65,025	519	94	41	56	9
261,121	713	93	41	63	9
1,046,529	1753	112	40	93	9

[2] O. Axelsson and P. S. Vassilevski. Algebraic multilevel preconditioning methods. II. *SIAM J. Numer. Anal.*, 27(6):1569–1590, 1990.

[3] O. Axelsson and P. S. Vassilevski. A black box generalized conjugate gradient solver with inner iterations and variable-step preconditioning. *SIAM J. Matrix Anal. Appl.*, 12(4):625–644, 1991.

[4] O. Axelsson and P. S. Vassilevski. Variable-step multilevel preconditioning methods. I. Selfadjoint and positive definite elliptic problems. *Numer. Linear Algebra Appl.*, 1(1):75–101, 1994.

[5] J. Bramble. *Multigrid methods*. Chapman & Hall/CRC, 1993.

[6] A. Brandt, S. McCormick, and J. Ruge. Algebraic multigrid (AMG) for sparse matrix equations. In *Sparsity and its applications (Loughborough, 1983)*, pages 257–284. Cambridge Univ. Press, Cambridge, 1985.

[7] M. Brezina, A. Cleary, R. Falgout, V. Henson, J. Jones, T. Manteuffel, S. McCormick, and J. Ruge. Algebraic multigrid based on element interpolation (amge). *SIAM Journal on Scientific Computing*, 22(5):1570–1592, 2001.

[8] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive smoothed aggregation (αSA) multigrid. *SIAM Rev.*, 47(2):317–346, 2005.

[9] W. Briggs and S. McCormick. *A multigrid tutorial*. Society for Industrial Mathematics, 2000.

[10] T. Chartier, R. D. Falgout, V. E. Henson, J. Jones, T. Manteuffel, S. McCormick, J. Ruge, and P. S. Vassilevski. Spectral AMGe (ρAMGe). *SIAM J. Sci. Comput.*, 25(1):1–26, 2003.

[11] G. H. Golub and Q. Ye. Inexact preconditioned conjugate gradient method with inner-outer iteration. *SIAM J. Sci. Comput.*, 21(4):1305–1320 (electronic), 1999/00.

[12] W. Hackbusch. *Multi-grid methods and applications*, volume 4. Springer Verlag, 1985.

[13] J. K. Kraus. An algebraic preconditioning method for M-matrices: linear versus non-linear multilevel iteration. *Numer. Linear Algebra Appl.*, 9(8):599–618, 2002.

[14] I. Lashuk and P. S. Vassilevski. On some versions of the element agglomeration AMGe method. *Numer. Linear Algebra Appl.*, 15(7):595–620, 2008.

[15] S. F. McCormick. Multigrid methods for variational problems: further results. *SIAM J. Numer. Anal.*, 21(2):255–263, 1984.

[16] S. F. McCormick. Multigrid methods for variational problems: general theory for the V-cycle. *SIAM J. Numer. Anal.*, 22(4):634–643, 1985.

[17] Y. Notay. Flexible conjugate gradients. *SIAM J. Sci. Comput.*, 22(4):1444–1460 (electronic), 2000.

[18] Y. Notay and P. S. Vassilevski. Recursive Krylov-based multigrid cycles. *Numer. Linear Algebra Appl.*, 15(5):473–487, 2008.

[19] J. W. Ruge and K. Stüben. Algebraic multigrid. In *Multigrid methods*, volume 3 of *Frontiers Appl. Math.*, pages 73–130. SIAM, Philadelphia, PA, 1987.

[20] Y. Saad. *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, second edition, 2003.

[21] V. V. Shaĭdurov. *Multigrid methods for finite elements*, volume 318 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995. Translated from the 1989 Russian original by N. B. Urusova and revised by the author.

[22] U. Trottenberg, C. Oosterlee, and A. Schüller. *Multigrid*. Academic Pr, 2001.

[23] P. Vaněk, J. Mandel, and M. Brezina. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing*, 56(3):179–196, 1996. International GAMM-Workshop on Multi-level Methods (Meisdorf, 1994).

[24] P. S. Vassilevski. Hybrid V-cycle algebraic multilevel preconditioners. *Math. Comp.*, 58(198):489–512, 1992.

[25] P. S. Vassilevski. *Multilevel block factorization preconditioners*. Springer, New York, 2008. Matrix-based analysis and algorithms for solving finite element equations.

[26] J. Xu. *Theory of multilevel methods*. PhD thesis, PhD thesis, Cornell University Ithaca, NY, 1989.

[27] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Rev.*, 34(4):581–613, 1992.

[28] J. Xu and L. Zikatanov. The method of alternating projections and the method of subspace corrections in Hilbert space. *J. Amer. Math. Soc.*, 15(3):573–597 (electronic), 2002.

[29] J. Xu and L. Zikatanov. On an energy minimizing basis for algebraic multigrid methods. *Comput. Vis. Sci.*, 7(3-4):121–127, 2004.

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA.

E-mail address: hu_x@math.psu.edu

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA 94551, U.S.A.

E-mail address: panayot@llnl.gov

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA.

E-mail address: xu@math.psu.edu